

# Implementing a Low-Cost Computer-Based Patient Record: A controlled vocabulary reduces database design complexity

Daniel J. Essin, MD LAC+USC Medical Center

Thomas L. Lincoln, MD LAC+USC Medical Center and RAND Corporation

*In order to build a computer-based patient record (CPR) system suitable for use in solo and small group practice settings it is necessary to use development methods that minimize cost. Design complexity is a major source of high cost. Reducing complexity should result in lower development, deployment and maintenance costs as well as higher reliability. We have developed a simplified relational model and have used that model, in conjunction with a controlled vocabulary, to implement a CPR that can capture and store patient examinations and other forms of clinical notes as well as laboratory and other test results. The information can be viewed in a familiar document format and it can be accessed for other types of processing using standard Structured Query Language (SQL) techniques. The database, as implemented, uses inexpensive components resulting in a system that is not prohibitively expensive for solo practitioners and small groups. In addition the architecture is scalable and can accommodate very large numbers of patients and practitioners.*

## INTRODUCTION

For many years it has been our objective to develop systems and applications that practicing physicians could use, in lieu of paper, to document their patient care activities and to provide a point of clinical reference so that treatment decisions would not have to be made "in the dark." It has been frustrating, at times, that all of the advances made in hospital-based information technology have not done much for the average office-based practitioner.

The assumption that computerized recordkeeping and data management will allow physicians to deliver high quality care at a lower cost is shared by many in Medical Informatics. Although this assumption is widely accepted, there are virtually no systems in wide-spread use, regardless of cost, that deliver the desired results. The most convincing demonstrations of how information technology can be applied to clinical care are found in large organizations whose systems are beyond the reach of the average practitioner. Recent changes in healthcare

policy and funding have shifted the emphasis toward out-patient care, adding to the urgency that information technology be developed that is suitable in scope and cost for the likely customers. Considering the amount of effort that has been expended, it is our conclusion that this problem cannot be solved by brute force; it needs to be redefined and/or divided before it can be conquered.

The primary sources of difficulty with the traditional approach to system development appears to be the sheer magnitude of the task. The traditional medical record contains hundreds of specialized forms. Some Hospital Information Systems (HIS) have identified over 25,000 data elements just to accommodate Admissions, Scheduling, Order Entry and basic Nursing Documentation. This facilitates specific retrieval and processing tasks but the complexity of the design typically requires a large development team and produces expensive systems. Much of the expense is associated with maintaining and modifying a system that was complex to start with. Were these systems to include physician's examination records and other forms of clinical documentation using the same approach, the number of data elements might well double and the complexity of the inter-element relationships, data integrity constraints, business rules would increase by an order of magnitude.

Other approaches have been proposed and novel systems have been built by various developers. Some have elected to store the clinical material, such as lab and x-ray reports and dictated clinical notes, in the form of text documents<sup>1,2</sup>. This approach greatly facilitates the display of individual documents or results for the clinician but requires powerful and sophisticated hardware and software to search the text effectively or to extract quantitative data from the documents for analytical processing. At the present time, small-scale versions of these systems have not been produced, effectively preventing these techniques from entering the domain of the small office or clinic.

The central issue is one of methodology. The "25,000 elements" approach requires extensive

analysis and planning to determine the "correct" design before the database is created. This is necessary because it is difficult to restructure an RDBMS while it is in use. The "text processing approach"<sup>3</sup> emphasizes the structural similarity of all documents, even though their content may vary. It is possible to add new documents types without the need to restructure the accumulated information. The fundamental difference in the two approaches is that the first assumes that "correctness" and design stability can be achieved, the latter assumes that there is no such thing as a "correct" design and that it would not remain stable even if there were.

In order to discuss cost-effectiveness, low cost and effectiveness must be defined. Among practices that use dictation extensively, the cost of transcription services approximates \$500/physician/month\*. This effectively limits the expenditure that most practices could justify to a range \$6,000 - \$18,000 every 3-5 years, unless there was added reimbursement for charting electronically. Low-reimbursement, primary care practices can be expected to favor the low end of the range. Effectiveness is defined as: 1) perception of usefulness, 2) ability of non-technical physicians to retrieve information, 3) data structured so that it can be retrieved either for viewing or for use by automated processes, 4) ability to operate on a small scale (single user), 5) ability to meet new needs by reconfiguration rather than reprogramming, 6) ability to scale-up in size or complexity, 7) no requirement to connect to an HIS, and 8) ability to exchange information with other systems.

It is our hypothesis that it is possible to produce a low-cost but effective CPR by combining commercial off-the-shelf components with an approach to information modeling and system construction that minimizes design complexity. The objective of this project was to complete the design and implementation a CPR database to interact with a pen-based application designed that would allow physicians and other healthcare professionals to document patient examinations and encounters.

## BACKGROUND

Apart from reducing the cost, there is theoretical justification for pursuing this simplification. A complex database design interferes with the full utilization

of the information that is accumulated. The semantics of the application's data model (as reflected in the table, column and relationship names), although familiar to the application developers, often remain a mystery to users. The semantics, although key to understanding the data and formulating queries, are often poorly documented and the user has no way to discover them by querying the system. Users rarely have permission to read the system tables and most RDBMS do not allow queries that would require performing a join on table or column names as output. The solution to this dilemma is to create a data model with fewer tables but one in which the tables include the formerly unqueryable table and column names as data. The full realization of this concept was described by Litwin et. al.<sup>5</sup> as First Order Normal Form (1ONF).

## METHODS

We used an iterative scenario-based form of knowledge engineering to define: 1) a series of outline templates that describe common medical events such as "H&P for Acute Illness in Child," and 2) a controlled vocabulary and the semantics needed to document each type of event. The goal of this activity is to provide coverage for more than 90-95% of the findings that are likely to occur during the event. The application uses the controlled vocabulary to determine which data elements are eligible for inclusion in the output to the CPR database. The users may include narrative comments in many places to add detail or to address an unanticipated situation.

The Controlled Vocabulary Source is constructed in a relational database consisting of five primary tables and defines a total of 58 data elements. The first is a table of Semantic Types and Relations drawn from the UMLS Knowledge Sources<sup>12</sup> with additional semantic subtypes. The second table defines sets of terms that are used together in the application. These include clusters of findings, treatments, historical elements, modifiers, etc. The third table contains detailed, application related data about each item in each set that is needed by the user-interface and the output engine. The fourth table defines any content modifiers specific to an item. This table specifies which units of measurement are to be used, which units are preferred, and what anatomic sites and/or methods are appropriate for the measurement in question.

---

\* Personal communication with several transcription vendors and physicians in private practice.

The final vocabulary table contains sets of records that define outline templates. The records in a set determine the semantic framework of a documentation session. They define which choices will be made available to the user and how the output will be organized and which entries are mandatory. With the exception of the context information, which is supplied programmatically, these outlines fully determine the appearance and behavior of the client application. An accessory table provides version control information about the outline template table.

The CPR database defines 45 data elements and is divided into three primary tables. The first table is the master event record. It is patterned after the HL7 OBR segment and incorporates many of the same elements. Key to the design is a text or memo field that can hold between 32 Kbytes and 2 gigabytes of text (depending on the RDBMS) and is fully searchable (not all RDBMS provide this capability). This insures that any information can be located by an appropriate query even if it is only recorded in a narrative addendum. The second table stores event details and is patterned after the HL7 OBX segment. One detail record is created for each pertinent finding. The third table stores, in separate records, all modifiers that were appended to the entry by the user. This provides more flexibility than an earlier iteration that was patterned after the Read Coding as implemented in MIQUEST<sup>13</sup> which allows for only 2 quantitative modifiers.

The database files were created in the Microsoft<sup>†</sup> Access format (\*.mdb). The client applications are written in Microsoft Visual Basic 3.0. Various clients have been used (from 25mhz-386SL<sup>‡</sup> to 40mhz-486 DX/2 processors). Servers have used either 50mhz-486DX or 75mhz-486 DX/4 processors. Client operating systems have been Windows 3.1 for Pen Computing, Windows for Workgroups 3.11 with Pen Extensions and Windows NT 3.5. Windows NT 3.5 and Windows for Workgroups 3.11 have been used as Server operating systems. Complete installations have been made on both wired and wireless local area networks and provides response times of 0.25 to 2 seconds depending on the network topology and the speed of the client and server processors. Performance has remained at these levels in databases containing 300+ megabytes of data and in tables with more than 375,000 records. The applica-

tions have been tested using Microsoft SQL Server as the RDBMS running on Windows NT 3.5 with essentially equivalent results.

## RESULTS

The client application produces a hybrid data structure using the outline templates and vocabulary as source information. Part of the output is an SGML-like structured document<sup>3</sup> that contains a copy of all information that was entered and is easily rendered for viewing. The other component is a set of traditional database records, each of which represents one significant finding or entry. Both components are inserted into the CPR database, presently housed in an RDBMS. In the event that the design of the relational database is altered, rows and columns that conform to the new design can be regenerated from the structured document without altering it in any way (unalterability is a requirement of a paperless medical record).

Using the data-structures and applications described here, an experienced user can define an outline template for a new purpose in 3-6 hours. The controlled vocabulary is organized into small components which have been proven to be highly reusable. As an example, a 2 page entry screen of perinatal history data that originally took 3 weeks to program was converted to an outline template in less than 2 hours. After using the new template, the users identified several fields that they had overlooked in their original design. The template was updated in 15 minutes without reprogramming and was again ready for use.

The information model for patient-clinician encounters specifies:

- 1) a set of context elements including:
  - a) what type of event it is
  - b) who is the patient
  - c) who performed the examination
  - d) where and when did the event occur
  - e) is the event linked to other events (visits)
  - f) is the event related to or an addendum to another event.
- 2) a set of event components, each of which records a facet of the event and includes:
  - a) semantic information, such as what type of component it is and how it is encoded
  - b) the descriptive and/or quantitative values of the component
  - c) additional elements to further characterize the component or assist in processing.

<sup>†</sup> All software names are trademarks of Microsoft Corp.

<sup>‡</sup> All processor names are trademarks of Intel Corp.

The output of the client application conforms to this information model<sup>6</sup> which defines various classes of information bearing objects<sup>7</sup>, each of which captures the information associated with a class of events. The event structure is similar to that described by Huff et. al.<sup>8,9</sup> and the event components are patterned after ASTM E1238 and the corresponding Chapter 7 of HL7<sup>10,11</sup>.

The following queries illustrate how the data can be retrieved from the CPR database for clinical use. These queries can be difficult to formulate using other conventions.

1) Get the data for a growth chart on Patient #1:  
`"SELECT quant, units, event_date FROM event_master, event_detail, event_detail_mods WHERE event_master.pid = 1 and event_detail.id = event_master.id and event_detail_mods.id = event_detail.id and result_name = 'weight' AND mod_type = 'result' ORDER BY event_date;"`

2) In which clinics has Patient #1 been seen and what kind of documentation was created: `"SELECT DISTINCT event_loc + ' ' + document_class AS exp1 FROM event_master WHERE pid = 1;"`

3) Which patients have received HIB vaccine in the past 3 months: `"SELECT lname, fname, medrecnum, birthdate FROM entity, event_master, event_detail WHERE result_group = 'PLAN' AND result_category = 'VACCINE' AND result_name = 'HIB' AND event_detail.id = event_master.id AND entity.id = event_master.pid AND event_date > Fmt( (DateVal ( Fmt(Now(), 'm/d/yyyy') )-90), 'm/d/yyyy' );"`

4) What prescriptions have been written for Patient #1 in the past month: `"SELECT event_date, result FROM event_master, event_detail WHERE event_master.pid = 1 AND event_detail.event_id = event_master.id AND event_date > Fmt( (DateVal ( Fmt( Now(), 'm/d/yyyy'))-30), 'm/d/yyyy') AND result_group = 'MEDS' ORDER BY event_date DESC;"`

## DISCUSSION

This project demonstrates that the acquisition cost of a CPR system can be as low as \$3,000 for a single user including the equipment need to run it. This means that small practices can afford to invest in electronic medical records. The approach that we have taken has produced an engine and applications whose behavior and properties, at any given time, are defined by the semantic data that resides in the

vocabulary source; changing the semantics changes the behavior.

No programming skills are required to configure the system or extend the vocabulary. Cost of ownership is minimized because most system maintenance and expansion can be performed by anyone who becomes familiar with the data structures and the semantics of the system and with off-the-shelf query and reporting tools.

Discipline is required to control and manage the process and this requires an investment of time and effort. Failure to organize the task can result in a problem similar to Keyword Drift<sup>14</sup>, a phenomenon whereby the semantics of an application wander over time. Keyword (perhaps better called Semantic) Drift occurs when users or developers who do not have good information about what terms are currently active continually invent new keywords and new rules for categorizing and indexing the same information that they have coded before.

Work in the area of controlled vocabulary offers mechanisms to realize stable and consistent application semantics. These efforts introduce a framework that can facilitate semantic browsing, determination of meaning, definition of relationships between terms, and maintenance<sup>15</sup>. It remains to be verified by subsequent evaluation, whether Semantic Drift occurs as more individuals elaborate the semantic framework and create additional outline templates. Presumably managing the semantics will prove to be less difficult than managing change in a traditional entity-relationship database.

The primary challenge in using any information system is dealing with changes and demands for new functionality that arise after people start using the system. This is due to the fact that every real project has to prioritize the requirements and begin somewhere. The controlled vocabulary that drives the applications described here is not something that can be abstractly determined to be "correct" any more than a database design can be "correct." It is primarily a device to meet the needs of practicing clinicians who must document, retrieve and process information about their actual experience with patients. These needs determine the content and organization of the vocabulary. Even more importantly, the demand for ongoing change is the result of learning to apply a new technology. As they gain experience, their vision, skills and expectations evolve; changing requirements and priorities are the natural by-product. Considering the emphasis on Continuous

Quality Improvement (CQI), a system that did not provoke this kind of response or could not participate in the CQI process might be deemed a failure.

A number of clinicians who have worked with the system have remarked that the outline templates are similar to the clinical pathways and practice guidelines that they are trying to implement. Since the outlines are user-definable, an immediate opportunity for cost-savings presents itself. Instead of buying a proprietary application to perform one of those tasks in isolation, a group of users can address a specific documentation or data gathering objective with tools they already know.

Although our design provides a mechanism to create more elaborate inter-record semantic linkages, we have not needed to do so at this time. Such links would be necessary to support automated inference but at present the primary outputs of the CPR are documents for human viewing and various queries based on the qualitative and quantitative components. Some inter-record linkages are provided by the RDBMS but primarily as referential integrity constraints. Embedded semantic linkages will become necessary when it becomes feasible to introduce an inference engine that can perform more elaborate associative processing.

## ACKNOWLEDGMENTS

The authors would like to thank Stanley Huff, MD and Roberto Rocha, MD of the University of Utah College of Medicine for their generosity in exchanging information about their work in progress.

## References

1. Vries JK, et. al.: Building a medical multimedia database system to integrate clinical information: an application of high-performance computing and communications technology. *Bulletin Of The Medical Library Association*, vol. 83(1): 57-64, 1995.
2. Culp KS, Myers DL, Keiser RD and Woods RD: The Scott & White Electronic Medical Record System (EMRS): Technologic Basis and Operational Characteristics, Scott & White Clinic and Hospital, Temple Texas, 1994.
3. Essin, DJ: Intelligent Processing of Loosely Structured Documents as a Strategy for Organizing Electronic Health Care Records. *Methods of Information in Medicine*, 32:265-68, 1993.
4. Essin DJ and Lincoln TL: "Pen Based Systems: Design Criteria for Event-Driven User Interfaces," AMIA Spring Meeting, San Francisco, May 1994.
5. Litwin W, Ketabchi M, Krishnamurthy R: "First Order Normal Form for Relational Databases and Multidatabases," *SIGMOD Record*, 20, 4, (December 1991), 74-76.
6. Essin, DJ and Lincoln, TL: "An Information Model for Medical Events" *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*, AMIA, (November 1994).
7. Guha RV and Lenat DB: Enabling Agents to Work Together, *Commun. ACM*, 37:7, (July 1994), 127-142.
8. Huff SM, Haug PJ, Stevens LE, Dupont RC and Pryor A: HELP The Next Generation: A new Client-Server Architecture. *Proceedings of the 18th Annual Symposium on Computer Applications in Medical Care*, AMIA, (November 1994).
9. Huff SM, Rocha RA, Bray BE, Warner HR, and Haug PJ: "An Event Model of Medical Information Representation", *JAMIA*, p 116-134, 1995.
10. ASTM Draft Standard E1238, rev 4.2.1, Standard Specification for Transferring Clinical Observations between Independent Computer Systems, ASTM, 1992.
11. Health Level Seven Standard Version 2.2: An Application Protocol for Electronic Data Exchange in Healthcare Environments. Health Level Seven, Inc., 1994.
12. UMLS Knowledge Sources: 5th Experimental Edition, National Library of Medicine, April, 1994.
13. MIQUEST and HQL: Morbidity Information Query and Export Syntax and Health Query Language. Analysis of Activities in General Practice Project, British NHS, April, 1993.
14. Essin, DJ and Lincoln, TL: "Healthcare Information Architecture: Elements of a New Paradigm" *reprinted in* *Proceedings of the National Computer Security Conference*, Baltimore, MD Oct. 1994.
15. Cimino, JJ, Representation of Clinical Laboratory Terminology in the Unified Medical Language System, *Proceedings of the 15th Annual Symposium on Computer Applications in Medical Care*, AMIA, (November 1991).